

Reduce to the MACs - Privacy Friendly Generic Probe Requests

Johanna Ansohn McDougall¹[0000-0001-7457-2181], Alessandro Brighente²[0000-0001-6138-2995], Anne Kunstmann¹[0009-0009-8990-9571], Niklas Zapatka¹[0009-0005-3841-1475], and Hannes Federrath¹

¹ University of Hamburg, Hamburg, Germany
{johanna.ansohn.mcdougall, firstname.lastname}@uni-hamburg.de

² University of Padova, Padua, Italy
alessandro.brighente@unipd.it

Abstract. Since the introduction of active discovery in Wi-Fi networks, users can be tracked via their probe requests. Although manufacturers typically try to conceal Media Access Control (MAC) addresses using MAC address randomisation, probe requests still contain Information Elements (IEs) that facilitate device identification. This paper introduces *generic probe requests*: By removing all unnecessary information from IEs, the requests become indistinguishable from one another, letting single devices disappear in the largest possible anonymity set.

Conducting a comprehensive evaluation, we demonstrate that a large IE set contained within undirected probe requests does not necessarily imply fast connection establishment. Furthermore, we show that minimising IEs to nothing but Supported Rates would enable 82.55% of the devices to share the same anonymity set. Our contributions provide a significant advancement in the pursuit of robust privacy solutions for wireless networks, paving the way for more user anonymity and less surveillance in wireless communication ecosystems.

Keywords: Probe Requests, Wi-Fi Tracking, Privacy-Preserving Technologies

1 Introduction

Wireless communication enables seamless connectivity and data exchange among devices. Particularly the ubiquitous use of active discovery has raised serious concerns regarding user privacy: Probe requests are actively sent by devices attempting to join a network, and the first step required for connection establishment between a user's device and an Access Point (AP). Albeit modern devices typically use MAC address randomisation and omit to send known Service Set Identifiers (SSIDs) unless searching for hidden networks [3], the information contained within the IE of the probe request still serves as a device fingerprint [10, 13, 17, 20, 22-25]. This fingerprint can be used to track device users, approximate the amount of people in a specified area and even trilaterate and thereby locate the signal origin up to an accuracy of 1.5 m [3].

In response to these limitations, this paper introduces *generic probe requests*: We propose the reduction of IE content to the bare minimum. We study the implications of generic probe requests for functionality, privacy, device identification, and tracking prevention. To this end, we contribute the following:

- We analyse the IE of probe requests and determine the minimal information required to receive probe responses. Our results show that the SSID and Supported Rates field are sufficient to receive valid probe responses.
- We analyse the impact of a reduced IE on functionality, privacy and security.
- We propose Generic Probe Requests and show that they provide a sufficiently large anonymity set, encompassing 82.55% of the probe requests. Thanks to the reduced content of their IEs, Generic Probe Requests resist correlation attacks undermining user privacy.

Our results show that Generic Probe Requests make devices as indistinguishable as possible, defending users from attacks targeting device fingerprinting via the IE content while simultaneously having no impact on the actual connection establishment, since they are only used during active discovery. While various publications have proposed to reduce IE content and randomise sequence numbers [17, 18, 25], to the best of our knowledge, we are the first to study the implications of minimising and unifying probe requests to the maximum.

This paper is structured as follows: In Section 2, we provide a background on MAC addresses, probe requests and connection establishment in Wi-Fi networks. Additionally, we introduce the Time-to-Traffic metric. We present Related Work in Section 3 and the attacker model in Section 4. Section 5 introduces the concept of generic probe requests and analyses its impact on functionality, privacy and security. Finally, Section 6 concludes the paper.

2 Background

This section gives a background on MAC addresses, probe requests and Wi-Fi connection establishment. In addition to that, we introduce the Time-to-Traffic metric that is used in our experiments.

2.1 MAC Addresses

A MAC address is a 48-bit network address used in Wi-Fi networks [16]. It serves to identify frame senders and receivers on the data link layer. Every Network Interface Controller (NIC) is assigned a fixed and globally unique Universally Administered Address (UAA) by its manufacturer. The first three bytes, the Organisationally Unique Identifier (OUI), identify the manufacturer. The last three bytes are assigned by the manufacturer. The UAA can be substituted by a Locally Administered Address (LAA). UAAs and LAAs are distinguished by the U/L-bit contained in every MAC address, which is the second-least significant bit of the most significant byte of a MAC address [15]. An example of LAAs are randomised MAC addresses, used to conceal the UAA while sending probe requests to prevent tracking via the MAC address.

2.2 Probe Requests

A mobile device can identify known networks via active or passive scanning. During passive scanning, APs broadcast beacon frames containing their SSID every 102.4 ms. Clients scanning the beacons can initiate the connection establishment upon recognising a known SSID [11]. The alternative, active scanning, is more commonly used because of its reduced overhead [11]. Here, clients broadcast *bursts* of probe requests: a set of bundled requests sent within a short time and transmitted over several channels [3]. Modern devices typically send *undirected* probe requests containing an empty SSID tag. An AP receiving such a probe request responds with a probe response containing its SSID. *Directed* probe requests contain SSIDs of known networks. They are required to locate hidden networks, which do not advertise via beacons and only respond to directed probe requests. Devices running outdated Operating Systems (OSes), or ones misconfigured by their users, may also include SSIDs in their probe requests [3].

The transmission of probe requests is initiated by the MAC sublayer management entity (MLME), which invokes layer management functions. The MLME initiates a scan via the MLME-SCAN.request primitive, which also determines the content of the probe request [16]. This content can be surveyed in the IE. While a large range of parameters *can* be transmitted via the IE [16, Section 6.3.3.2.2], devices typically transmit a select few IE tags. These can include, but are not limited to the Supported Rates, Extended Supported Rates and the transmission channel (DS Parameter). Other common fields are High Throughput (HT), Very High Throughput (VHT) and High Efficiency (HE) Capabilities, advertising support for the IEEE 802.11n, 802.11ac and 802.11ax standard respectively. Extended Capabilities declare support for additional features beyond HT and VHT capabilities. The Interworking field enables seamless connectivity in heterogeneous network environments (IEEE 802.11u). Another common field contains vendor specific information.

2.3 Connection Establishment in Wi-Fi Networks

Wireless local area networks (WLANs) are based on the IEEE 802.11 standard [16]. Each WLAN is identified by its SSID, and mobile devices can identify known networks via network discovery (cf. Section 2.2). Subsequent connection establishment with an AP is done in several steps: first, IEEE 802.11 authentication is performed to grant a client access to the network. This encompasses two authentication and two association frames, in which the client and AP are paired and communication parameters and standard extensions negotiated. A successful association enables data transfer of frames of higher layers. Subsequently, WPA, WPA2 and WPA3 use Robust Security Network Association (RSNA), a suite of protocols for authentication and encryption. Upon successful completion, WLAN access is granted, and encrypted data frames can be exchanged.

While the term WLAN describes the standard, the Wi-Fi Alliance instead promotes the use of the term Wi-Fi, a trademark protecting all products certified to their Wi-Fi interoperability [27]. Since the term Wi-Fi is commonly used in anglophone publications, we adhere to this de-facto standard in this paper.

No.	Time	Source	Destination	Channel	Antenna	sign.	Info
3	312.535887684	3e:6a:81:c4:65:ab	Broadcast	-94	dBm		Probe Request, SN=18, FN=0, Flags=....., SSID=wildcard (Broadca...
4	312.537618051	D-LinkIn_32:91:68	3e:6a:81:c4:65:ab	7	-78	dBm	Probe Response, SN=171, FN=0, Flags=....., BI=100, SSID="hobbit...
5	312.644201889	3e:6a:81:c4:65:ab	Broadcast	-86	dBm		Probe Request, SN=20, FN=0, Flags=....., SSID=wildcard (Broadca...
6	312.645900029	D-LinkIn_32:91:68	3e:6a:81:c4:65:ab	7	-78	dBm	Probe Response, SN=174, FN=0, Flags=....., BI=100, SSID="hobbit...
7	315.004753259	D-LinkIn_31:c1:fc	Broadcast	-46	dBm		Probe Request, SN=0, FN=0, Flags=....., SSID=wildcard (Broadcas...
8	315.895378110	D-LinkIn_31:c1:fc	Broadcast	-46	dBm		Probe Request, SN=1, FN=0, Flags=....., SSID=wildcard (Broadcas...
9	315.806150359	D-LinkIn_31:c1:fc	Broadcast	-46	dBm		Probe Request, SN=2, FN=0, Flags=....., SSID="hobbitholes"
10	315.807884387	D-LinkIn_32:91:68	D-LinkIn_31:c1:fc	7	-78	dBm	Probe Response, SN=206, FN=0, Flags=....., BI=100, SSID="hobbit...
11	315.809829957	D-LinkIn_32:91:68	D-LinkIn_31:c1:fc	7	-78	dBm	Probe Response, SN=206, FN=0, Flags=...R..., BI=100, SSID="hobbit...
12	316.109813705	D-LinkIn_32:91:68	D-LinkIn_31:c1:fc	-78	dBm		Authentication, SN=213, FN=0, Flags=.....
13	316.11955273	D-LinkIn_31:c1:fc	D-LinkIn_32:91:68	-46	dBm		Association Request, SN=5, FN=0, Flags=....., SSID="hobbitholes"
14	316.112064573	D-LinkIn_32:91:68	D-LinkIn_31:c1:fc	-78	dBm		Association Response, SN=214, FN=0, Flags=.....
15	316.119766244	D-LinkIn_32:91:68	D-LinkIn_31:c1:fc	-78	dBm		Key (Message 1 of 4)
16	316.122974140	D-LinkIn_31:c1:fc	D-LinkIn_32:91:68	-46	dBm		Key (Message 2 of 4)
17	316.125309557	D-LinkIn_32:91:68	D-LinkIn_31:c1:fc	-78	dBm		Key (Message 3 of 4)
18	316.127114332	D-LinkIn_31:c1:fc	D-LinkIn_32:91:68	-46	dBm		Key (Message 4 of 4)
19	316.137851901	D-LinkIn_31:c1:fc	Broadcast	-46	dBm		Data, SN=8, FN=0, Flags=p.....T

Fig. 1: Measurement points in the TtT-metric. Packet 6 is the starting point, the first probe response to the last probe request sent via a randomised address. The endpoint of the metric is the beginning of data transmission in packet 19.

2.4 Time-to-Traffic

We introduce the Time-to-Traffic (TtT) metric for measuring the duration of the connection process. With it, we measure the time between the last probe response received before the MAC address changes to the LAA, and the first data frame exchanged after connection establishment. Its duration can be observed in Fig. 1: Packet 6, marked in blue, is the starting point of the TtT. The endpoint of the measurement is the first data frame, packet 19. This can be explained as follows: a device probing for nearby networks usually uses a randomised MAC address. The probe requests in this stage are typically undirected and therefore do not contain an SSID. Once a device identifies a known network via a probe response, it switches its MAC address to its UAA or a fixed LAA per network. Typically, further probes are transmitted via this address, sometimes first undirected and then directed, sometimes either of the two. Subsequently, connection establishment is initiated. Some devices mostly omit these directed or undirected probe requests using the LAA, and immediately initialise the connection after receiving a probe response from a known network. To receive comparable results, we therefore choose to measure the TtT beginning with the last probe response to the randomised MAC address.

3 Related Work

To track a mobile device over a longer period of time, it is necessary to apply de-randomisation techniques to probe requests to correlate the bursts originating from different MAC addresses with a single device. The research on this topic can be split into two competing perspectives: attacks, and countermeasures.

3.1 Attacks

Linking bursts to devices requires a unique pattern in the requests, e.g. the globally unique UAA [5]. When a device uses a regularly changing LAA, other fingerprinting attacks are possible. The main focus then lies in fingerprinting

using IEs and sequence numbers. After the theoretical risk was noticed [10], a first empirical assessment was provided [25]. This attack was improved by taking the MAC address into account [17]. Subsequently, other IE fields were shown to be usable for fingerprinting devices [13, 14, 21]. Other approaches to defeat MAC address randomisation model probe request frame association in a flow network and use minimum-cost flow optimisation [23], or combine IE fingerprinting with clustering approaches [20, 24].

An additional source of information for device fingerprinting are timing attacks: Pattern in request transmission times were shown to be specific to the device driver [9]. This attack was subsequently extended to single out individual devices [18]. Using a timing attack, it is possible to identify a device’s hardware by measuring channel switch times during scanning [11].

3.2 Countermeasures

Albeit fingerprint techniques are improved continuously, manufacturers are slow to deploy countermeasures. The solutions are mainly academic, e.g. to rely solely on passive discovery, which decelerates the connection establishment by only 0.6 seconds [26]. This can be improved by accelerating passive scanning [11]. While this is a valid approach in most cases, hidden networks can only be reached via directed probe requests. By transmitting the SSID as a hash, combined with the current MAC address and sequence number of the packet, the use of hidden networks is possible in a privacy friendly manner [3].

Improvements to active scanning include the use of MAC address randomisation, avoiding unique MAC addresses by using a regularly changing, randomised LAA [12]. An alternative is MAC address masquerading, where the own UAA is replaced by the MAC address of a nearby device [9]. Fingerprinting could be further hampered by minimising transmitted IEs and randomising sequence numbers [17, 25], and introducing randomness into the sending pattern of single packets and bursts [18]. The use of cryptographic measures for content protection of probe requests would create an overhead of 2 seconds for the key exchange and 0.5 seconds for the transmitted probe request [13]. To assess the countermeasures deployed in mobile devices, a comparison of 160 mobile devices manufactured between 2012 and 2020 revealed that MAC address randomisation has been widely adopted, and sequence number randomisation is emerging. However, IE fingerprinting is still possible [8].

Our proposal to employ generic probe requests enhances these countermeasures: The IE minimisation maximise the anonymity set size and ensures that tracking via the content of probe requests is impeded.

4 Attacker Model

We consider a passive attacker. Their objective is to track the movements of individuals within a public area, such as a campus or shopping centre. This is

accomplished by eavesdropping on the probe request bursts emitted by smartphones during active scanning. The adversary has installed a sufficient number of Wi-Fi receivers in the area to triangulate their targets' positions.

We assume that the attacker can link probe request bursts to individual devices by exploiting unique patterns, e.g. the UAA. If a device uses MAC address randomisation, a pattern may still be derived by fingerprinting the associated IEs. If two probe requests exhibit the same patterns, the attacker assumes that they originate from the same device. We assume the attacker is thereby capable of associating probe requests to their sending devices and can track single users.

5 Generic Probe Requests

Various attacks to track devices despite MAC address randomisation utilise IE-fingerprinting techniques [13, 17, 20, 24, 25]. Our approach to counter them hence lies in reducing the content of the IE to the bare minimum. To gauge how far the content of a probe request can be reduced, we transmit probe requests via a USB Wi-Fi adaptor using Scapy³, an interactive program for packet manipulation in Python. Two more Wi-Fi adaptors are used to monitor the traffic via Wireshark on the non-overlapping channels 1 and 6 of the 2.4 GHz bands, which are the most frequently used channels in the vicinity of our experimental setup. The script used to send probe requests can be found on GitHub⁴.

Observation 1: By removing more and more fields from the IE of the probe request, we determine that the SSID and Supported Rates field are the only ones necessary for probe requests. Neither of these fields is required to contain legitimate information in order to allow for appropriate reception of probe responses.

Since the content of the IE is typically used for tracking, we propose to reduce it to the bare minimum established in this experiment. To evaluate the feasibility of this approach, we subsequently evaluate the time required for connection establishment depending on the content of the IE to determine whether a large IE causes faster connection establishment.

5.1 IE Content Analysis

In theory, IEs in Probe Requests serve clients to communicate their capabilities to nearby APs. This information helps the APs to respond appropriately and efficiently, ensuring that the client can connect to the network in a manner aligning with its requirements and the network's capabilities. To estimate whether transmitting capabilities using a complex IE field accelerates the connection establishment, we compare the connection establishment times of five different

³ <https://scapy.net/>

⁴ <https://Github.com/heddha/scapy-probe-requests>

Table 1: Comparison of the Information Element and the respective Time-to-Traffic of probe requests of five different devices.

	Supported Rates (bytes)	Ext. Supported Rates (bytes)	DS Parameter Set (bytes)	HT Capabilities (bytes)	Vendor Specific (bytes)	Extended Capabilities (bytes)	HE Capabilities (bytes)	Interworking (bytes)	Frame Size undirected (bytes)	Frame Size directed (bytes)	Average TtT (seconds)	Median	Std. Deviation
DWA-171	8	4	-	-	-	-	-	-	68	79	3.71	3.69	0.16
Intel 8265	8	4	1	26	7	-	-	-	108	277	1.92	1.67	0.73
iPhone SE	4	8	1	26	-	8	27	-	139	184	2.77	2.41	0.98
iPad Pro	4	8	1	26	28	8	-	7	152	163	2.54	2.8	1.17
Raspberry Pi	4	8	1	26	131	-	-	-	236	247	2.73	1.86	2.26

devices and correlate them with the size of their IE. The devices used in our tests are a mobile phone (iPhone SE 2020 running iOS 17), a Wi-Fi dongle (D-Link DWA-171 Nano Wi-Fi USB Adaptor), a laptop running an Intel AC 8265 wireless card, a single-board computer (Raspberry Pi Model 3B+), and a tablet (iPad Pro running iPadOS 15). They represent typical household appliances.

The tests are conducted as follows: We connect our devices to an access point, which we turn on at non-fixed intervals. We measure how long it takes the devices to reconnect to the AP. To measure whether a complex information element causes quicker connection establishment, we utilise the TtT metric introduced in Section 2.4. We conduct the tests 7 times per device.

The results of the experiment can be seen in Table 1. All devices transmit a minimum of Supported Rates and Extended Supported rates. The probe requests of the *DWA-171 adaptor* have a total frame size of 68 bytes, 12 bytes of which made up by Supported Rates and Extended Supported Rates. The TtT required is 3.71 seconds on average. The *Intel 8265* laptop transmits undirected probe requests with a total frame size of 108 bytes. The TtT is 1.92 seconds on average, and the IE additionally contains channel information, HT Capabilities, and vendor-specific information. The frame size of the *iPhone* is 139 bytes, and its TtT 2.77 seconds. In addition to the channel and HT Capabilities, it transmits 35 bytes of Extended Capabilities and HE Capabilities. The *iPad Pro* transmits 152 bytes long probe request frames and has a TtT of 2.54 seconds. The largest IE body is contained in the undirected probe requests of the *Raspberry Pi*, with 236 bytes and a TtT of 2.73 seconds.

As this comparison shows, the Intel 8265 card is by far the fastest despite the second-smallest frame size. This is also represented by its median of 1.67. The Intel 8265’s standard deviation is the second smallest of all with 0.73. The iPad follows as the second-fastest, however, its standard deviation of 1.17 is higher than both the iPhone or the Raspberry Pi (with the largest frame body). The DWA-171 antenna is the slowest to establish a connection and also has the smallest frame size. With a median of 3.69, it has the smallest standard deviation of only 0.16. The results show that there does not appear to be a direct correlation between the amount of information transmitted in undirected

probe requests and the speed of the connection establishment after the device discovery. The underlying reasons for fast connection establishment can likely be found in optimised implementations, and the very slow connection establishment of the DWA-171 might be attributed to the slow data transmission via a USB 2.0 Type A wireless dongle.

On the other hand, Table 1 also allows for a comparison of the frame sizes of *directed* probe requests: Directly before connection establishment between client and AP as well as after the client has switched to the LAA, it often sends one or more directed probe requests. A directed probe request from the DWA-171, the iPad, and the Raspberry Pi contains no additional information besides the SSID. The iPhone transmits 184 instead of 139 bytes in its directed probe request, with the additional bytes concerning three vendor-specific tags and the SSID. The Intel 8265 laptop transmits a directed probe request of 277 bytes, with the additional information concerning Extended Capabilities, Mesh ID, FILS request parameters, and three vendor-specific tags containing information on WPS, P2P, and multi-band operation. While this does not necessarily have to correlate, comparing the connection establishment with respect to the IE size of directed probe requests could be interesting in future work.

5.2 Proposition: Minimisation of IE content

In order to not impede any significance that directed probe requests might have on communication establishment, we propose *generic probe requests, in which the IE content of undirected probe requests is reduced to the bare minimum*: By modifying the IE to only contain the Supported Rates and (empty) SSID fields, user privacy can be increased immensely, while simultaneously satisfying only the minimum requirements for legitimate IE content. The modifications only have to be applied to undirected probe requests as the majority of transmitted probe requests and are the core element of the research on deanonymisation of probe requests. Protecting them is essential to assure user privacy. The reason for limiting the protection to undirected probe requests lies in the nature of *directed* probe requests, which are only transmitted in three scenarios:

- (i) Directly before connection establishment,
- (ii) during an existing connection to ensure the connection is maintained with the AP with the strongest signal and
- (iii) while searching for hidden networks.

In the first two cases, the directed probe requests are sent using the LAA, which is typically either stable for an extended period of time per network, or unchanging [3]. The transmission of the LAA and that of the SSID serve as a good enough identifier to enable trivial device tracking, which makes the protection of IE content irrelevant in this scenario. Limiting the content of undirected probe requests has no impact on connection establishment or the connection itself: Both adhere to standards that do not rely on probe requests, using the established methods explained in Section 2.3.

In the following, we investigate the impact a reduced IE content has on the exchange of capabilities with the AP.

5.3 Impact on Functionality

Connection establishment as described in Section 2.3 requires four stages: (1) active or passive device discovery, (2) authentication, (3) association and (4) RSNA.

Observation 2: Connection establishment is also possible when using passive discovery only. Since passive discovery results in a connection just as efficient and stable as when using active discovery [11,26], the capabilities that define the connection must be exchanged in another manner.

When studying the IEEE Wi-Fi standard [16, Fig. 4-30], it becomes apparent that when the AP sends probe responses, as well as when the client sends the IEEE Std. 802.11 association request, "security parameters" are transmitted. A comparison of the fields contained in association requests [16, Table 9-34] and probe requests [16, Table 9-38] reveals that association requests and probe requests contain many overlapping fields, but while association requests contain 46 elements, probe requests contain only 34 elements. Out of these 34 elements, 15 are not present in association requests. They can be observed in Table 2.

Table 2: The fields present in probe requests, but *not* in association requests, and their existence or substitution by other fields in association requests and probe responses.

Field	AP-CSN	Change Sequence	Channel Usage	Cluster Probe	DSSS Parameters	Est. Service Param. Inb.	Est. Service Param. Outb.	Extended Request	FILS Request Parameters	Mesh ID	PV1 Probe Response Option Request	S1G Relay Discovery	Vendor Specific Req. SSID List
In Probe Req.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
In Probe Resp.	✓	✓	✓	†	✓	✓	✓	(✓)	✓	†	✓	†	
In Assoc. Req.								(✓)		†	(✓)		†

(✓): Fields further negotiating the specific tag are present
 †: Substituted by a different field

When looking at association requests, they contain elements negotiating FILS session activation and the use of S1G Relays. Therefore, their use must have been declared elsewhere when using passive discovery. An investigation of probe *responses* confirms this: of the 15 elements not present in association requests, 8 are transmitted in probe responses. In addition to 84 other fields, probe responses (and also beacons) contain a FILS Indication field, which explains why devices using passive discovery can be aware of FILS capabilities without having requested information on it.

The fields which are exclusively present in probe requests are Request, Extended Request, and Vendor Specific Request, SSID List, FILS Request Parameters, PV1 Probe Response Option and Cluster Probe. The Cluster Probe element

as well as the FILS Request Parameters are covered instead by the Extended Cluster Report field, respective FILS indication in probe responses and subsequent FILS parameters in the association frame. The SSID List element bundles multiple SSIDs that the device can connect to. In practice, requesting multiple SSIDs is instead done by sending separate probe requests for each SSID. The PV1 Probe Response Option [16, Tables 9-305 - 9-310] field is a bitmap of capabilities and compatibility, containing requests to respond with the capabilities the AP has. Most fields that can be requested via this bitmap are optionally present in probe responses and association requests. While this appears to be a useful field, connection establishment is possible, as well as efficient in passive discovery, without either the PV1 Probe Response Option or the Request, Extended Request, and Vendor Specific Request tags. Therefore, these fields appear to be unnecessary in the exchange of capabilities.

In fact, to test the variability in probe responses with respect to the probe requests they respond to, we monitor probe responses in three different settings: in response to (1) scripted probe requests as used in Section 5, containing only an empty SSID field and Supported Rates, (2) those of the Intel 8265, and (3) those of a Raspberry Pi, with the longest IE field of all tested devices.

Observation 3: Regardless of the IE content of probe requests, the AP always responds with the same capabilities.

To confirm this, we observe the probe responses of 10 different APs in the vicinity of our setup over an extended period of time, spanning probe responses to various passersby and a large range of devices, with the same result. We conjecture that even probe requests containing a reduced IE content will receive all necessary capabilities of the AP via its probe responses. Since all relevant information in probe requests are also exchanged via the combination of probe response and association request, and the IE content does influence probe responses, we conjecture that reducing the IE content to the bare minimum is feasible, and would reduce complexity and remove redundancy from wireless communication.

To estimate the improvement that generic probe requests have on the anonymity of single users, we subsequently analyse and compare anonymity sets.

5.4 Impact on Privacy: Anonymity Set Determination

An anonymity set defines the number of users sharing the same identifiers, thus making them indistinguishable from one another. To estimate the privacy gain of generic probe requests, we evaluate the anonymity set size with reduced IE content like Vanhoef et al. in 2016 [25]. We use a subset of the Sapienza dataset [4], containing probe requests recorded at a train station in 2013, containing 376117 probe requests. MAC address randomisation was first introduced in iOS 8 in 2014 [1], and in Android 6.0 [2] in 2015. The dataset therefore contains *very few probe requests that employ some form of MAC address protection*. It is perfect for the anonymity set determination: To perform the same analysis on more recent

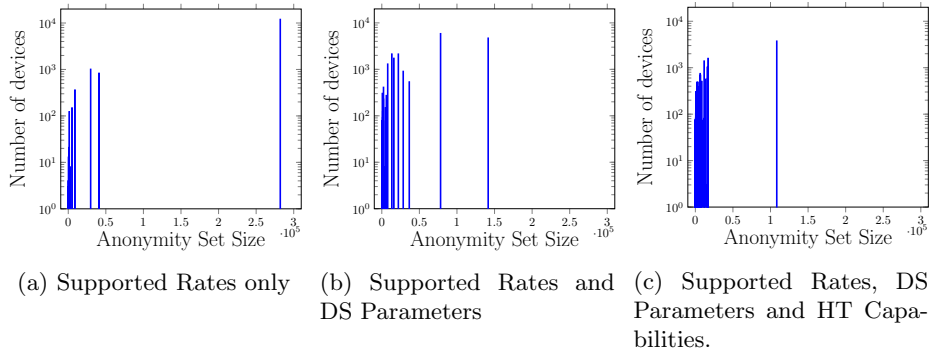


Fig. 2: The amount of devices contained in the same anonymity set when the IE content is reduced to contain (a) only Supported Rates, (b) Supported Rates and DS Parameters, and (c) Supported Rates, DS Parameters and HT Capabilities.

data, we would first have to apply de-randomisation strategies to the data set, e.g. define an attack that maps probe requests to single devices. This has been done in several publications (cf. Section 3.1), but is out of scope for this paper.

Using the Wireshark filter `!(wlan.sa[0] & 0x02)`, we determine which probe requests originate from globally unique MAC addresses via the U/L bit (cf. Section 2), and find that 374736 are UAAs. These make up the pruned subset used for the subsequent evaluation. These probe requests originate from 14622 distinct MAC addresses. Since the dataset originates from before the introduction of MAC address randomisation, it is safe to assume that *the number of distinct MAC addresses correlates with the number of devices*.

The results of our evaluation can be seen in Fig. 2a: The graph shows that reduction of the IE to contain only Supported Rates results in 19 distinct anonymity sets. Around 17.45% of the probe requests are in 18 smaller anonymity groups containing one to about 1000 devices. The remaining group unifies 82.55% of the devices in one large anonymity set. All of these devices share the same Supported Rates 2, 4, 11, and 22. Fig. 2b shows a reduction of the content of the IE to Supported Rates and DS Parameters. This reduction results in 61 anonymity sets, the largest of which now contains 5920 devices, which amounts to 40.49%. In 225508 cases, the DS Parameters were not contained within the probe request. In the remaining cases, the DS Parameter set contained channels 1, 2, 11, or 12.

Figure Fig. 2c shows the anonymity set size in case the IE contains both Supported Rates, DS Parameters, and HT capabilities. The distribution is spread out significantly over 276 sets and the largest anonymity set contains 25.78%, or 3770, of the devices. These results show that the less information is contained within probe requests, the larger the anonymity set most devices are contained in. It is therefore necessary to reduce and unify IE content as much as possible. To determine how much exactly, we regard prior attacks and the exact IE elements they use to fingerprint devices in the following.

Table 3: A comparison of the fields used for IE fingerprinting. Publications that focused on certain fields with high entropy are highlighted in grey.

Authors	Publication Year	MAC Address	SSID	Supported Rates	Ext. Supp. Rates	DS Parameter Set	HT Capabilities	VHT Capabilities	Ext. Capabilities	Vendor Specific	all other fields	IE Field Sequence	BSS Membership	Transm. Frequency	Sequence No.	RSS
Pang et al. [19]	'07	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-
Cunche et al. [6]	'14	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-
Freudiger [10]	'15	✓	-	-	-	-	-	-	-	-	-	-	-	-	✓	-
Vanhoef et al. [25]	'16	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	-
Robyns et al. [21]	'17	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	✓	-	-
Zhao et al. [28]	'19	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-
Dagelić et al. [7]	'19	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-
Gu et al [13]	'20	-	-	✓	✓	✓	✓	✓	-	✓	-	-	-	-	-	-
Uras et al. [24]	'20	-	-	-	✓	✓	✓	✓	-	-	-	✓	-	-	-	-
Tan et al. [23]	'21	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	✓	✓
Pintor et al. [20]	'22	-	-	-	-	✓	✓	-	✓	✓	-	-	-	-	-	-
He et al. [14]	'23	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	✓

5.5 Impact on Security: Resistance to IE-Concerned Attacks

To gauge the impact that a reduction of the IE content has on the security, a comparison of the fields used in IE-concerned attacks is performed in Table 3. The table highlights the fields particularly regarded in this publication, including the Supported Rates, the DS Parameter Set and the HT Capabilities in blue, purple and green. Additionally, the rows containing publications that selected specific fields due to their distinguishing features and high entropy are highlighted in grey. The conclusion that becomes apparent in this visual comparison is that *no publication that selected fields depending on high entropy chose to include the Supported Rates*. The publications that included the Supported Rates incorporated them not due to high entropy, but to maximise the fingerprint. This shows that the Supported Rates play only a minor role in identifying devices. This knowledge, in combination with our calculation of the anonymity set size of reduced IE content, show that *removing all tags but the Supported Rates and the empty SSID field prevents existing attacks*. This way, users are protected from device tracking via the IE contained in their probe requests, while ensuring that probe responses can still be received.

6 Conclusion

Reducing the content of probe requests to the minimum while enlarging the anonymity set to the maximum has been evaluated from different perspectives: Our experiments showed, that the only IE tags necessary for probe requests to receive probe responses are the Supported Rates and SSID, both of which can be empty. We argue to unify the content of probe requests by transmitting only Supported Rates and the SSID. We furthermore showed that devices with probe requests containing the largest IE set do not necessarily correspond to the fastest connection establishment. On the other hand, while the very large size of *directed*

probe requests sent by the Intel 8265 and its fast connection establishment do not necessarily have to correlate, it still poses the question whether large IEs in *undirected* probe requests serve a purpose at all, or whether a combination of generalised and very slim undirected probe requests and, upon identifying a known network, sending information-rich directed probe requests might improve both the privacy, as well as the connection establishment speed.

We argue that other factors than the size of the IE must be the cause for the efficiency of the connection establishment, and that for the sake of increased user privacy, reducing the IE content in *undirected* probe requests to the minimum is a valid option. To evaluate the anonymity provided by such minimisation, we calculate the anonymity set size of the Sapienza train station data set and find out, that reducing IE content to contain only Supported Rates allows for 82.55% of the devices to be contained in the same anonymity set. This is a significant discovery that would protect a large number of users with very little effort.

Altogether, generic probe requests as proposed in this paper inhibit attacks targeting the IE by reducing its content to the bare minimum and thereby protect user privacy.

References

1. Aaron Mamiit: Apple implements random MAC address on iOS 8. Goodbye, marketers. <https://www.techtimes.com/articles/8233/20140612/apple-implements-random-mac-address-on-ios-8-goodbye-marketers.htm> (2014)
2. Android Developers: Android 6.0 Changes. <https://developer.android.com/about/versions/marshmallow/android-6.0-changes> (2023)
3. Ansohn McDougall, J., Burkert, C., Demmler, D., Schwarz, M., Hubbe, V., Federath, H.: Probing for Passwords – Privacy Implications of SSIDs in Probe Requests. In: Ateniese, G., Venturi, D. (eds.) Applied Cryptography and Network Security. pp. 376–395. Springer International Publishing, Cham (2022)
4. Barbera, M.V., Epasto, A., Mei, A., Kosta, S., Perta, V.C., Stefa, J.: Crawdad sapienza/probe-requests (2022). <https://doi.org/10.15783/C76C7Z>
5. Bernardos, C.J., Zuniga, J.C., O’Hanlon, P.: Wi-Fi internet connectivity and privacy: Hiding your tracks on the wireless Internet. In: IEEE Conference on Standards for Communications and Networking (CSCN). pp. 193–198. IEEE (2015)
6. Cunche, M., Kaafar, M.A., Boreli, R.: Linking Wireless Devices Using Information Contained in Wi-Fi Probe Requests. *Pervasive and Mobile Computing* **11**, 56–69 (2014)
7. Dagelić, A., Perković, T., Čagalj, M.: Location Privacy and Changes in WiFi Probe Request Based Connection Protocols Usage Through Years. In: International Conference on Smart and Sustainable Technologies (SpliTech). pp. 1–5. IEEE (2019)
8. Fenske, E., Brown, D., Martin, J., Mayberry, T., Ryan, P., Rye, E.C.: Three Years Later: A Study of MAC Address Randomization In Mobile Devices And When It Succeeds. *PETS’2021* pp. 164 – 181 (2021)
9. Franklin, J., McCoy, D., Tabriz, P., Neagoe, V., Van Randwyk, J., Sicker, D.: Passive data link layer 802.11 wireless device driver fingerprinting. In: *USENIX Security’06*. USENIX Association (2006)
10. Freudiger, J.: How Talkative is Your Mobile Device? An Experimental Study of Wi-Fi Probe Requests. In: *WiSec’15*. ACM (2015)

11. Goovaerts, F., Acar, G., Galvez, R., Piessens, F., Vanhoef, M.: Improving Privacy Through Fast Passive Wi-Fi Scanning. In: *Secure IT Systems*. pp. 37–52. Springer (2019)
12. Gruteser, M., Grunwald, D.: Enhancing location privacy in wireless lan through disposable interface identifiers: A quantitative analysis. In: *Proceedings of the 1st ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots*. p. 46–55. WMASH '03, Association for Computing Machinery, New York, NY, USA (2003)
13. Gu, X., Wu, W., Gu, X., Ling, Z., Yang, M., Song, A.: Probe Request Based Device Identification Attack and Defense. *Sensors* **20**(16) (2020)
14. He, T., Tan, J., Chan, S.H.G.: Self-supervised association of wi-fi probe requests under mac address randomization. *IEEE Transactions on Mobile Computing* **22**, 7044–7056 (2023)
15. IEEE: IEEE Std 802-2014 - Local and Metropolitan Area Networks: Overview and Architecture. <https://ieeexplore.ieee.org/document/6847097> (2014)
16. IEEE: IEEE Std 802.11 - Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications (2020)
17. Martin, J., Mayberry, T., Donahue, C., Foppe, L., Brown, L., Riggins, C., Rye, E.C., Brown, D.: A Study of MAC Address Randomization in Mobile Devices and When It Fails. *PETS'2017* **4**, 268–286 (2017)
18. Matte, C., Cunche, M., Rousseau, F., Vanhoef, M.: Defeating MAC Address Randomization Through Timing Attacks. In: *WiSec'16*. pp. 15–20. ACM (2016)
19. Pang, J., Greenstein, B., Gummadi, R., Seshan, S., Wetherall, D.: 802.11 User Fingerprinting. In: *MobiCom'07*. p. 99. ACM (2007)
20. Pintor, L., Atzori, L.: Analysis of wi-fi probe requests towards information element fingerprinting. In: *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*. pp. 3857–3862 (2022)
21. Robyns, P., Bonné, B., Quax, P., Lamotte, W.: Noncooperative 802.11 mac layer fingerprinting and tracking of mobile devices. *Secur. Commun. Networks* **2017**, 6235484:1–6235484:21 (2017)
22. Rütermann, T., Benabbas, A., Nicklas, D.: Know thy quality: Assessment of device detection by wifi signals. In: *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. pp. 639–644. IEEE (2019)
23. Tan, J., Chan, S.H.G.: Efficient Association of Wi-Fi Probe Requests under MAC Address Randomization. In: *INFOCOM'21*. pp. 1–10. IEEE (2021)
24. Uras, M., Cossu, R., Ferrara, E., Bagdasar, O., Liotta, A., Atzori, L.: Wifi probes sniffing: an artificial intelligence based approach for mac addresses de-randomization. In: *2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)* (10 2020)
25. Vanhoef, M., Matte, C., Cunche, M., Cardoso, L.S., Piessens, F.: Why MAC Address Randomization is Not Enough: An Analysis of Wi-Fi Network Discovery Mechanisms. In: *Asia CCS '16*. p. 413–424. ACM (2016)
26. Waltari, O., Kangasharju, J.: Quantifying the information leak in ieee 802.11 network discovery. In: Chowdhury, K.R., Di Felice, M., Matta, I., Sheng, B. (eds.) *Wired/Wireless Internet Communications*. pp. 207–218. Springer International Publishing, Cham (2018)
27. wikipedia: Wi-Fi. <https://en.wikipedia.org/wiki/Wi-Fi> (2023)
28. Zhao, F., Shi, W., Gan, Y., Peng, Z., Luo, X.: A localization and tracking scheme for target gangs based on big data of Wi-Fi locations. *Cluster Computing* **22**(1), 1679–1690 (2019)